

Sławomir Sapanowski

Okręgowa Komisja Egzaminacyjna w Łodzi

Oprogramowanie WEKA – przykład analizy wyników z zastosowaniem data minig

WEKA is not a bird.

Wpisując hasło WEKA do wyszukiwarki internetowej, uzyskujemy następujący opis (Wikipedia): **Weka** (*Gallirallus australis*) - gatunek nietlonego ptaka z rodziny chruścielowatych (Rallidae). Występuje endemicznie w Nowej Zelandii.



Cóż ten miły ptak ma wspólnego z analizą wyników testowania? Otóż WEKA (Waikato Environment for Knowledge Analysis) jest również nazwą pakietu oprogramowania, które od kilku lat jest rozwijane i udoskonalane na Uniwersytecie Waikato w Hamilton (Nowa Zelandia). Jest to system umożliwiający używanie różnych metod badawczych od sieci neuronowych i bayesowskich aż do zadań opartych o *data minig*¹. Podobne zalety oferuje pakiet *Statistica*, SPSS czy (przy odrobinie wprawy) poczciwy Excell. Po co zatem jeszcze jeden program, który *de facto* wykonuje te same analizy i obsługi, którego będziemy musieli się od nowa uczyć? Odpowiedź jest prosta. Cena wymienionych programów jest na tyle wysoka, że może skutecznie zniechęcić najbardziej wytrwałego amatora dociekań statystycznych. WEKA nie ma tej wady – rozpowszechniana jest jako oprogramowanie *open source* na licencji GNU General Public License. Oznacza to, że jest zupełnie darmowa i możemy korzystać z niej w dowolnym zakresie.

Po zainstalowaniu i uruchomieniu programu ukazuje się niezbyt imponujący ekran z głównym menu, który jednak kryje w sobie ogromne bogactwo treści.

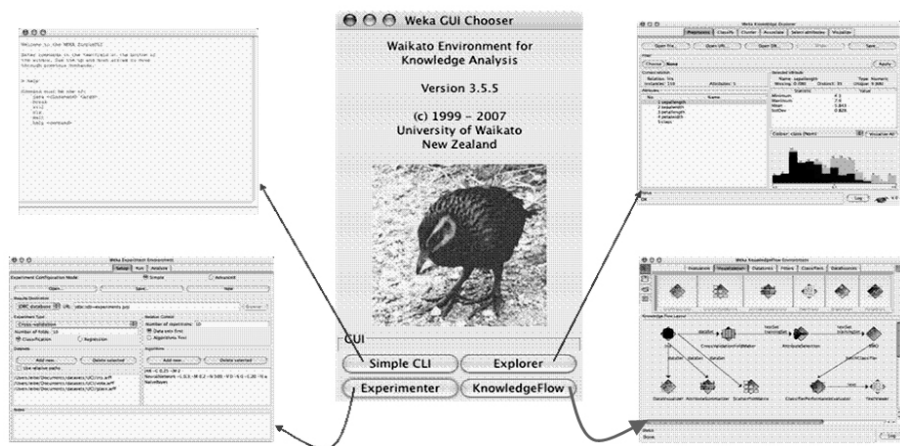
Simple CLI – uruchamiamy środowisko typu „wiersz poleceń” (raczej dla zaawansowanych użytkowników);

Explorer – środowisko, w którym możemy analizować dane;

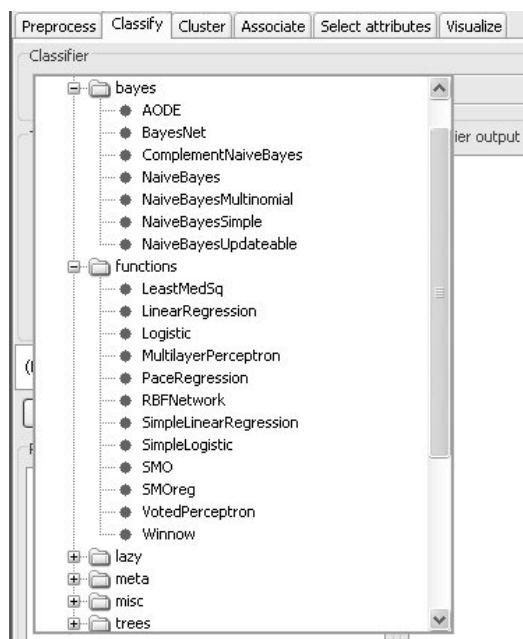
Experimenter – panel pozwalający na projektowanie i testowanie hipotez statystycznych;

KnowledgeFlow - kolejna część WEKI, która w zasadzie zawiera funkcjonalność „Explorera” i jest oparta o technologię drag-and-drop.

¹ Data mining (zglębianie danych). Data mining jako proces analityczny przeznaczony jest do eksploracji dużych zbiorów danych (zazwyczaj odnoszących się do zjawisk gospodarczych lub rynkowych) w poszukiwaniu reguł i systematycznych zależności pomiędzy zmiennymi, a następnie do oceny wyników poprzez zastosowanie wykrytych prawidłowości do nowych podzbiorów danych. Ostatecznym celem data mining jest przewidywanie. (<http://www.statsoft.pl/textbook/glosfra.html>)



Pełny przekaz o funkcjonalności programu wykracza poza ramy tego artykułu², dlatego też opis ograniczę do zakładki Explorer, która zawiera wiele narzędzi statystycznych nazywanych „filtrami”. Ich użycie jest stosunkowo proste (z reguły wystarcza kilka kliknięć), a różnorodność ogromna – od prostych regresji liniowych przez regresję logistyczną (IRT!) aż do tak skomplikowanych procesów, jak sieci neuronowe.



² Wszystkich zainteresowanych odsyłam do obszernego opisu możliwości programu WEKA, który można znaleźć pod adresem <http://www.cs.waikato.ac.nz/ml/weka/>.

WEKA odczytuje dane z plików .csv, C4.5 oraz w specjalnym formacie .arff. Możliwe jest również importowanie danych z internetu lub bazy danych SQL. Program pracuje nie tylko z danymi liczbowymi i nominalnymi, ale również z tekstem (wystarczy w pliku z danymi zadeklarować odpowiedni typ danych).

Po raz pierwszy pakiet WEKA wykorzystałem do oszacowania tzw. „efektu egzaminatora” na sprawdzianie w szóstej klasie szkoły podstawowej (2009) (por. Zapieraczyński, Sapanowski, Zbidi, 2007). Na podstawie odpowiedzi ucznia udzielonych na zadania zamknięte, płci oraz informacji o dysleksji chciałem przewidzieć oczekiwany wynik za zadania otwarte (sprawdzane przez egzaminatora). Przygotowane dane (w formacie .arff) podałem analizie za pomocą prostego filtra (LinearRegression) – regresja liniowa.

Linear Regression Model

Zad_otw =

```

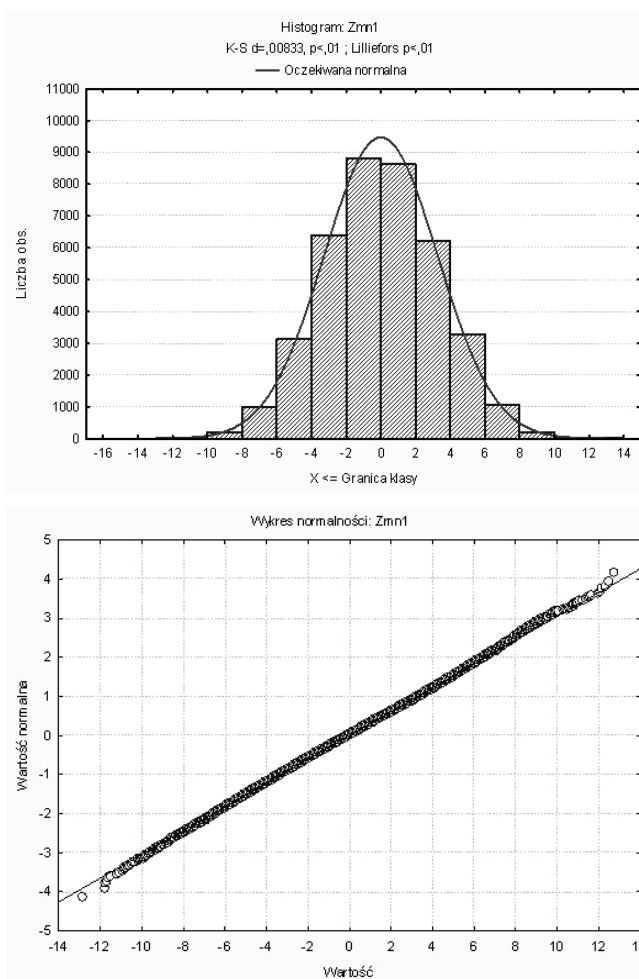
1.7751 * plec=K +
1.4447 * dysl=T +
0.8675 * zad1 +
0.6186 * zad2 +
0.6192 * zad3 +
0.8705 * zad4 +
0.8273 * zad5 +
0.6727 * zad6 +
0.6317 * zad7 +
0.556 * zad8 +
1.1694 * zad9 +
1.1225 * zad10 +
1.0309 * zad11 +
0.5747 * zad12 +
0.7618 * zad13 +
0.518 * zad14 +
0.7957 * zad15 +
0.7624 * zad16 +
0.9861 * zad17 +
0.9791 * zad18 +
1.2564 * zad19 +
0.6955 * zad20 +
-2.8954
    
```

Time taken to build model: 1.49 seconds

Otrzymane wyniki są zaskakujące. Współczynnik o wartości 1,7751 występujący przy płci ucznia wskazuje, że dziewczęta o podobnym wyniku jak chłopcy w strefie zadań zamkniętych (podobnych umiejętnościach) uzyskiwały średnio o 1,77 punktu więcej w sferze zadań otwartych. Podobny efekt występuje w grupie dyslektyków – uczeń ze specyficznymi trudnościami w nauce, ze względu na łagodniejsze kryteria oceniania, zyskuje przeciętnie 1,44 punktu.

Różne wartości współczynników przy zadaniach zamkniętych wskazują na to, że estymacja uzyskanych przez zdającego punktów za zadania otwarte nie może opierać się tylko i wyłącznie na sumie punktów za WW. Wyraźnie widać, że niektóre zadania są „ważniejsze” niż inne. Dla uczniów mających taki sam wynik za zadania WW będziemy przewidywać różną liczbę punktów za zadania otwarte.

Naturalnie rodzi się pytanie o jakość takiego przewidywania. Najistotniejszym kryterium poprawności modelu jest to, czy rozkład reszt³ jest rozkładem normalnym. Możemy to sprawdzić na podstawie histogramu i wykresu normalności zamieszczonych poniżej.



³ Przez resztę rozumiemy różnicę pomiędzy wynikiem uzyskanym a wynikiem przewidywanym w części dotyczącej zadań otwartych.

Szczególnie wykres normalności wskazuje na poprawność zastosowanej metody, a co za tym idzie możemy wykorzystać ją do szacowania tzw. „efektu egzaminatora”.

Obliczając średnią reszt dla grupy egzaminatorów, chcielibyśmy, aby nie różniła się ona znacznie od średniej dla całej populacji⁴. Otrzymana liczba w pewien sposób charakteryzuje egzaminatora czy też grupę egzaminatorów. Jeżeli jest ujemna, oznacza to, że oceniający prace uczniów był bardziej surowy od innych, jeżeli dodatnia – łagodniejszy.

Oczywiście, interpretując ten wskaźnik, należy zachowywać dużą ostrożność w wyciąganiu wniosków⁵. Osoby, które analizowały wskaźniki EWD (obie metody opierają się na podobnej idei) doskonale wiedzą o pułapkach czyhających na niedoświadczonego analityka. Mimo to chciałbym zachęcić, szczególnie koordynatorów sprawdzianu i egzaminu gimnazjalnego, do analizy pracy grup egzaminatorów pod kątem jakości oceniania. Właściwe wykorzystanie opisanej metody może wpłynąć na usprawnienie pracy przewodniczących zespołów egzaminacyjnych z egzaminatorami. Daje również gwarancję na podwyższenie jakości sprawdzania, a także odbiera argumenty przeciwnikom egzaminów zewnętrznych, którzy podważają obiektywizm i rzetelność sprawdzania.

Bibliografia:

1. Dokumentacja programu WEKA, http://www.cs.waikato.ac.nz/ml/weka/index_documentation.html
2. Zapieraczyński, M. Sapanowski, S. Zbidi, A. *Zależność wyników uzyskiwanych przez uczniów z zadań otwartych od liczby punktów zyskanych za zadania zamknięte*, [w:] XIII KDE *Uczenie się i egzamin w oczach uczniów*, Łomża 2007.

⁴ Oczywiście średnia reszt dla całej populacji wynosi 0 i nie występuje korelacja między resztą a wynikiem z zadań WW.

⁵ Dla jasności treści artykułu pominąłem wątek przedziałów ufności dla średnich, które przy analizie „efektu egzaminatora” mają ogromne znaczenie.